

Article

The Impact of Global Sensitivities and Design Measures in Model-Based Optimal Experimental Design

René Schenkendorf ^{1,2,*} , Xiangzhong Xie ^{1,2,3}, Moritz Rehbein ^{2,4}, Stephan Scholl ^{2,4} and Ulrike Krewer ^{1,2} 

¹ Institute of Energy and Process Systems Engineering, TU Braunschweig, Franz-Liszt-Straße 35, 38106 Braunschweig, Germany; x.xie@tu-braunschweig.de (X.X.); u.krewer@tu-braunschweig.de (U.K.)

² Center of Pharmaceutical Engineering (PVZ), TU Braunschweig, Franz-Liszt-Straße 35a, 38106 Braunschweig, Germany

³ International Max Planck Research School (IMPRS) for Advanced Methods in Process and Systems Engineering, Sandtorstraße 1, 39106 Magdeburg, Germany

⁴ Institute for Chemical and Thermal Process Engineering, TU Braunschweig, Langer Kamp 7, 38106 Braunschweig, Germany; moritz.rehbein@tu-braunschweig.de (M.R.); s.scholl@tu-braunschweig.de (S.S.)

* Correspondence: r.schenkendorf@tu-braunschweig.de; Tel.: +49-531-391-65601

† Current address: Institute of Energy and Process Systems Engineering, TU Braunschweig, Franz-Liszt-Straße 35, 38106 Braunschweig, Germany.

Received: 14 February 2018; Accepted: 19 March 2018; Published: 21 March 2018



Abstract: In the field of chemical engineering, mathematical models have been proven to be an indispensable tool for process analysis, process design, and condition monitoring. To gain the most benefit from model-based approaches, the implemented mathematical models have to be based on sound principles, and they need to be calibrated to the process under study with suitable model parameter estimates. Often, the model parameters identified by experimental data, however, pose severe uncertainties leading to incorrect or biased inferences. This applies in particular in the field of pharmaceutical manufacturing, where usually the measurement data are limited in quantity and quality when analyzing novel active pharmaceutical ingredients. Optimally designed experiments, in turn, aim to increase the quality of the gathered data in the most efficient way. Any improvement in data quality results in more precise parameter estimates and more reliable model candidates. The applied methods for parameter sensitivity analyses and design criteria are crucial for the effectiveness of the optimal experimental design. In this work, different design measures based on global parameter sensitivities are critically compared with state-of-the-art concepts that follow simplifying linearization principles. The efficient implementation of the proposed sensitivity measures is explicitly addressed to be applicable to complex chemical engineering problems of practical relevance. As a case study, the homogeneous synthesis of 3,4-dihydro-1H-1-benzazepine-2,5-dione, a scaffold for the preparation of various protein kinase inhibitors, is analyzed followed by a more complex model of biochemical reactions. In both studies, the model-based optimal experimental design benefits from global parameter sensitivities combined with proper design measures.

Keywords: optimal experimental design; global parameter sensitivities; optimal design measures; robustification; point estimate method

1. Introduction

Mathematical models have been proven to be an indispensable tool in chemical engineering research and design. For instance, model-based and computer-aided concepts are the standard in process analysis [1–3], process design [4–6], and condition monitoring [7–9]. To gain the most benefit of model-based approaches, the model-building process itself becomes a crucial step in computer-aided process analysis and design. When it comes to mathematical modeling and model-based design, model developers are challenged by the problem of providing meaningful results. The situation is the same for the computer-aided process design of pharmaceutical processes and active pharmaceutical ingredients (API) syntheses [10,11]. In addition to a proper model structure determining the interconnection of involved quantities and species, the most critical part is to identify related model parameters [12,13]. When assuming a correct model structure, any mismatch of measurement data and simulation results is mainly attributed to biased model parameters. Optimal tuning of those model parameters according to the experimental data is an essential step in the process of model development where model parameters are iteratively adapted until simulation results fit the given measurement data best [12,14]. Thus, the quality of model-based results depends significantly on the quality of the parameter estimates. The parameter quality itself, however, depends on the quality of the measurement data and the conducted experiments, respectively. The operating conditions of the experiment indirectly determine the sensitivity of the model parameters to the simulation results. The parameters with a high sensitivity are likely to be reliably identified. Model parameters with low sensitivity, in turn, can be changed by an order of magnitude with little to no effect. Any attempt at a manual or algorithmic tuning of such insensitive parameters is error prone by definition. To avoid this kind of ill-posed parameter identification problem, model-based optimal experimental design (MB-OED) comes into play [14–17]. The aim of MB-OED is to redesign the experimental setup ensuring high sensitivities of all parameters over the course of the experimental run. To this end, a dynamic optimization problem has to be solved, which typically includes the following generic steps; see also Figure 1 for their interaction. First, the parameter sensitivities of a given model candidate have to be quantified. Based on the parameter sensitivities, a suitable design measure is defined and evaluated, which translates parameter sensitivities into algorithmically evaluable quantities. After specifying potential design variables of the experimental setup, which are practically feasible and allow the control of the main process states, an optimizer routine adapts the experimental setup iteratively. The resulting optimal design variables determine the setup of the next experimental run, which is expected to produce new informative measurement data. These data, in turn, represent operating conditions at which, in general, the model parameters show an increased sensitivity leading to more precise parameter estimates. This MB-OED loop is reiterated until the parameter estimates fulfill the needs of particular modeling tasks. While following the proposed work flow, i.e., (1) sensitivity quantification, (2) design measure evaluation, (3) dynamic optimization, and (4) running an optimally designed experiment, the final result depends critically on the sensitivity measure and its proper implementation. Any improper choice at this point will lead to sub-optimal experimental designs and non-informative data. Thus, reliable sensitivity measures are mandatory for MB-OED to provide the most informative data possible. This aspect is particularly relevant for syntheses of active pharmaceutical ingredients, where typically the number of novel drugs and the number of experimental runs are limited.

In the literature, most implementations of MB-OED are based on local sensitivities as part of the Fisher information matrix (FIM) [12,14,16,18–20]. Technically, the inverse of the FIM is used to approximate the covariance matrix of the parameter estimates [12,14,18]. This local approach assumes a linear relation between parameter perturbations and model responses—at least locally in the neighborhood of the correct reference parameter values. As the actual parameter values, however, are typically unknown, their current best estimates have to be inserted instead. Thus, the calculated local sensitivities may fall short as they do not describe the underlying nonlinearities and because of biased reference parameters [16,21]. To compensate for biased references, sensitivity gradients can be evaluated at various but representative parameter values within the given parameter domain

following a Bayesian experimental design principle [22–24]. Here, the basic idea is to get an improved approximation of the local sensitivities based on an averaged measure [25]. Nonlinear effects, however, still remain unaddressed. Most parameter identification problems of practical relevance in the field of pharmaceutical processes are characterized by nonlinear impacts of model parameters on model responses [19]. Thus, local sensitivities, in general, might lead to oversimplified inferences, less robust experimental designs, and an unnecessarily high number of experimental reiterations [21]. This last point, as mentioned previously, is of critical relevance when analyzing novel but exceedingly rare drug candidates.

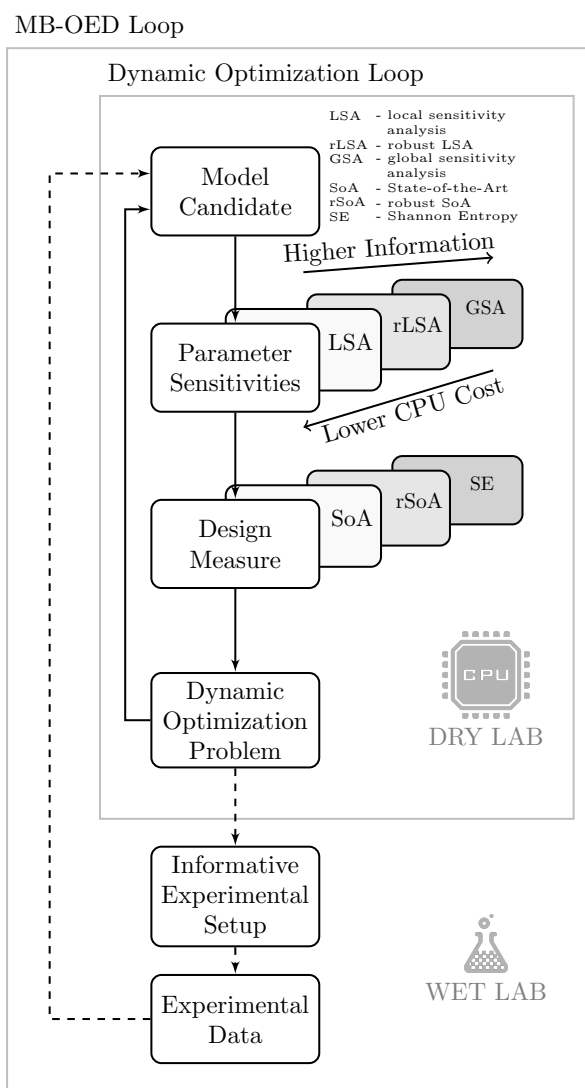


Figure 1. Key elements of the model-based optimal experimental design (MB-OED): the parameter sensitivity approach and the implemented design measure determine the quality of the designed experiment significantly.

Alternatively, in the last decade, the implementation of global sensitivities for MB-OED has been introduced in the literature [25–27]. Global sensitivities by definition take into account nonlinear effects and multivariate parameter dependencies. To the best of the authors' knowledge, the usefulness of GSA and its practical implementation for pharmaceutical and (bio)chemical processes have not been analyzed thus far. This manuscript shows how global sensitivities can be determined efficiently by a deterministic sampling rule even for complex models with a large number of model parameters, which makes GSA available for advanced MD-OED strategies.

The focus of this study is to compare and assess different parameter sensitivity measures, as well as an analysis of the need for novel design criteria for MB-OED. In particular, local sensitivity analysis (LSA), robust local sensitivity analysis (rLSA), and global sensitivity analysis (GSA) are introduced. In addition to state-of-the-art (SoA) design criteria, further alternatives are discussed in the following. Thus, the paper unfolds as follows. First, the basics of parameter sensitivity analysis as a key element of optimal experimental design are presented in Section 2. Here, the focus is also on the efficient implementation of global parameter sensitivities. Based on the derived sensitivity measures, optimal design metrics are derived in Section 3. In Section 4, the proposed concept is illustrated by two pharmaceutically relevant processes, the homogeneous synthesis of 3,4-dihydro-1H-1-benzazepine-2,5-dione, a scaffold for the preparation of various protein kinase inhibitors, and a more complex biochemical synthesis example. The conclusions are given in Section 5.

2. Sensitivity Measures

In the literature, different methods and concepts for parameter sensitivity analysis (SA) exist, ranging from local to global approaches [28–30]. In the context of MD-OED, sensitivity analysis is an inherent part of the optimization problem aiming to provide an optimal range of sensitivities over the course of the experimental run. In the following, we consider nonlinear state-space models of the form

$$\frac{dx(t)}{dt} = f(x(t), u(t), \theta), \quad (1)$$

$$y(t) = h(x(t), u(t), \theta), \quad (2)$$

where $x(t)$ is the vector of model states ($x \in \mathbb{R}^n$), $u(t)$ is the vector of system inputs ($u \in \mathbb{R}^r$), θ is the vector of parameters ($\theta \in \mathbb{R}^p$), and $y(t)$ is the vector of model outputs ($y \in \mathbb{R}^m$). Typically, only a minor subset of the states $x(t)$ can be measured directly or indirectly; i.e., $m < n$. The proposed concepts can be extended to more general model types, e.g., differential algebraic or partial differential equation systems, but are discussed here solely for nonlinear state space models for the sake of clarity. As the model parameters θ are typically unknown, they have to be identified by minimizing the error between simulation results, $y(t)$, and measurement data, $y^{\text{data}}(t)$. By following a maximum likelihood estimation procedure and assuming additive white measurement noise with a standard deviation of one [14], the parameter identification problem simplifies to

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^K ||y^{\text{data}}(t_i) - y(t_i)||^2, \quad (3)$$

where K is the number of measurement time points. According to the measurement noise and the Doob–Dynkin lemma [31], the identified model parameters $\hat{\theta}$ are random variables as well; for example, measurement uncertainties induce stochastic parameter uncertainties. The resulting parameter uncertainties, in turn, are determined by the data quality and parameter sensitivities alike.

2.1. Local Parameter Sensitivities

In practice, local sensitivities are the standard for analyzing the effect of parameter variations on model-based results [28,29,32]. Local sensitivities S^{LSA} are typically expressed by partial derivatives of the j -th model output, $y[j](t)$, with respect to the i -th model parameters, $\theta[i]$, as

$$S^{\text{LSA}}[i, j](t) = \frac{\partial y[j](t)}{\partial \theta[i]} \quad (4)$$

and summarized by the sensitivity matrix $S^{\text{LSA}}(t) \in \mathbb{R}^{m \times p}$ while $j = 1, \dots, m$ and $i = 1, \dots, p$. In the field of dynamical systems, the governing equations of the original model (Equation (2)) have to be extended by sensitivity terms of the states, S_x^{LSA} , which are solved in parallel according to

$$\dot{S}_x^{\text{LSA}} = \frac{\partial f}{\partial x} \cdot S_x^{\text{LSA}} + \frac{\partial f}{\partial \theta}; \quad S_x^{\text{LSA}}(t=0) = 0_{n \times p}. \quad (5)$$

Thus, the output sensitivities can be expressed as

$$S^{\text{LSA}} = \frac{\partial h}{\partial x} \cdot S_x^{\text{LSA}} + \frac{\partial h}{\partial \theta}. \quad (6)$$

Alternatively, the gradients (Equation (4)) can be calculated with various numerical methods, e.g., automatic differentiation or finite differences. Here, the interested reader is referred to [32,33] and references therein.

Technically, local sensitivities provide information about the relevance of model parameters in the neighborhood of a reference point within the defined parameter space, a reference which is typically unknown for most practical problems. In an ideal world, hypothetically, the reference point corresponds to the exact model parameters. Moreover, the general idea of local SA measures is based on linearization principles; i.e., the original, nonlinear model (Equation (2)) is linearized at the (unknown) reference point. Thus, nonlinear effects are ignored although the analyzed models may have strong nonlinearities involved [21]. The resulting approximation errors may render the outcome of such a sensitivity analysis meaningless and difficult to interpret in general. To compensate for the dependency of unknown reference points, a multi-point averaging approach can be applied as described in the following sections.

2.2. Robustification: Multi-Point Averaging Approach

As the reference points for the sensitivity gradients are typically unknown, local sensitivities may provide artifacts; i.e., insensitive parameters are identified as sensitive or vice versa. To avoid this kind of mis-classification, an averaged version of the sensitivity matrix (Equation (4)) seems to be a more robust measure [25,34] following a Bayesian experimental design principle [22–24] and is defined as

$$S^{\text{rLSA}}[i, j](t) = E[S^{\text{LSA}}[i, j](t)] = \int_{\Omega} S^{\text{LSA}}[i, j](t) pdf_{\theta[i]} d\theta[i], \quad (7)$$

where $E[\cdot]$ is the expected value, Ω is the support domain, and the probability density function (PDF), $pdf_{\theta[i]}$, represents the assumed parameter variation of the i -th parameter. According to Equation (7), for linear problems, the proposed concept simplifies to the local SA expression [25]. In most practical scenarios, there exists no analytical solution of Equation (7). Thus, Monte Carlo (MC) simulations are evaluated instead to solve a discretized approximation resulting in a multi-point averaging approach equal to

$$S^{\text{rLSA}}[i, j](t) = E[S^{\text{LSA}}[i, j](t)] \approx \sum_{k=1}^N S^{\text{LSA}}[i, j](\theta_k[i], t), \quad (8)$$

where $\theta_k[i]$ represents the k -th realization of the i -th parameter according to the given PDF. The sample number, N , determines the total number of function calls of Equation (5) needed to approximate the averaged sensitivity measure, i.e., for each parameter realization, θ_k , Equation (5) has to be vectorized and solved in parallel with the dynamic model (Equation (1)) that is also evaluated for θ_k . By using S^{rLSA} instead of the standard local quantities S^{LSA} , the resulting parameter sensitivities are more representative of nonlinear problems [34–36] and may provide a reliable base for subsequent MB-OED studies. As standard MC simulations might become prohibitive in terms of computational load, efficient sample-based approaches are mandatory and will be discussed in more detail in Section 2.4.

The multivariate interaction, however, of the analyzed model parameters is still not taken into account appropriately by the multi-point averaging concept. Here, the framework of global sensitivities seems to be more suitable to rank vaguely known parameters and to quantify their interactions.

2.3. Global Parameter Sensitivities

Global sensitivity measures treat parameters, θ , and model outcomes, $y(t)$, as random variables and aim to quantify the amount of variance that each parameter, $\theta[i]$, adds to the total variance of the output, $\sigma^2(y(t))$ [29,37,38]. In detail, the parameter ranking is done by the amount of output variance that disappears when the i -th parameter $\theta[i]$ is assumed to be known, $\sigma^2(\theta[i]) = 0$. For this particular parameter, a conditional variance, $\sigma_{-i}^2(y(t)|\theta[i])$, can be derived. Here, the subscript $-i$ indicates that the variance is taken over all parameters other than $\theta[i]$. As $\theta[i]$, in reality, is a random variable, the expected value of the resulting conditional variance, $E_i \left[\sigma_{-i}^2(y(t)|\theta[i]) \right]$, has to be analyzed. The subscript notation of E_i indicates that the expected value is taken only over the parameter $\theta[i]$ itself. Finally, the total output variance, $\sigma^2(y(t))$, can be split into two additive terms [29] equal to

$$\sigma^2(y(t)) = \sigma_i^2(E_{-i}[y(t)|\theta[i]]) + E_i[\sigma_{-i}^2(y(t)|\theta[i])]. \quad (9)$$

Here, the variance of the conditional expectation, $\sigma_i^2(E_{-i}[y(t)|\theta[i]])$, represents the contribution of parameter $\theta[i]$ to the total variance $\sigma^2(y(t))$. The normalized expression given in Equation (10) is known as the first-order Sobol sensitivity index [29] and is used in the following to analyze parameter sensitivities and multivariate parameter interactions, respectively:

$$S^{\text{GSA}}[i](t) = \frac{\sigma_i^2(E_{-i}[y(t)|\theta[i]])}{\sigma^2(y(t))}. \quad (10)$$

Similar to the multi-point averaging approach (Section 2.2), the integral terms associated with $\sigma^2(y(t))$, $E_{-i}[y(t)|\theta[i]]$, and $\sigma_{-i}^2(y(t)|\theta[i])$ are commonly evaluated with MC simulations [39]. For global sensitivity studies of complex models, MC simulations have prohibitive computational costs. Thus, while implementing an MB-OED strategy based on global sensitivities, highly efficient methods have to be used. The most relevant concepts in this direction are outlined in the next section.

2.4. Implementation Aspects

Over the last two decades, various methods have been developed to calculate Sobol indices efficiently. The most frequently used approaches are based on advanced MC simulations and polynomial chaos expansion (PCE) principles [40]. In comparison to ordinary MC simulations, advanced sampling concepts aim for a better convergence rate by an improved space-filling sampling strategy, i.e., to avoid clustering effects, as well as gaps in the p -dimensional sampling space [41]. Please note that p represents the dimension of the random vector, i.e., the number of not perfectly known model parameters. For instance, so-called quasi-MC methods have a convergence rate of $\mathcal{O}((\log N)^p/N)$, which for low-dimensional problems can lead to a significant reduction in computational costs in comparison to $\mathcal{O}(1/\sqrt{N})$ for ordinary MC [41]. On the other side, PCE-based approaches aim to replace the original CPU-intensive model with a handy and easy-to-evaluate surrogate model first. Then, some characteristics of the derived PCE models are used to express global sensitivities analytically by using the PCE coefficients directly or to run additional MC simulations at low computational costs [42]. The parameterization of the PCE model, however, requires a significant number of reference simulations of the original model. Both concepts, advanced MC methods and PCE, might be prohibitive for MB-OED problems because of their computational costs. Highly efficient uncertainty handling methods are needed instead. Here, the point estimate method (PEM) provides a fair compromise in terms of CPU load and accuracy [43,44].

The fundamental idea of the PEM is to choose sample points, θ_k , and associated weights, w_k , in accordance with the first central moments of given random variables [43]. However, in comparison

to MC simulations, these sample points are generated deterministically and not randomly. Assuming a three-dimensional parameter problem, for instance, the resulting sample set reads as

$$GF(\pm\vartheta) = \begin{bmatrix} \theta[1] \\ \theta[2] \\ \theta[3] \end{bmatrix} + \left\{ \begin{bmatrix} \vartheta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -\vartheta \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \vartheta \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -\vartheta \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vartheta \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -\vartheta \end{bmatrix} \right\}. \quad (11)$$

The generator function, $GF(\pm\vartheta)$, creates the sample set by permuting an element $\theta[i]$ of the original random vector θ one by one by the amount of ϑ . Thus, the scaling parameter, ϑ , controls the spread of the sample points in the p -dimensional parameter space. To calculate the expected value, the discretized integration problem reads as

$$E[g(\theta)] = \int_{\Omega} g(\theta) p d f_{\theta} d\theta \approx w_0 g(\theta) + w_1 g(GF(\pm\vartheta)), \quad (12)$$

with $w_0 = 1 - p/\vartheta^2$ and $w_1 = 1/2\vartheta^2$ assuming a standard Gaussian distribution. Moreover, the overall precision of the PEM can be increased gradually by considering higher-order central moments of θ . The more precise approximation scheme [44,45] results in

$$E[g(\theta)] = \int_{\Omega} g(\theta) p d f_{\theta} d\theta \approx w_0 g(\theta) + w_1 g(GF(\pm\vartheta)) + w_2 g(GF(\pm\vartheta, \pm\vartheta)), \quad (13)$$

where $GF(\pm\vartheta, \pm\vartheta)$ implies the simultaneous variation of two elements of the given parameter vector θ . Thus, the number of generated sample points, θ_k , scales quadratically with $2p^2 + 1$ for a p -dimensional parameter problem: a good balance of computational load and approximation power.

The four unknown coefficients, ϑ , w_0 , w_1 , and w_2 , can be expressed analytically as

$$\vartheta = \sqrt{\frac{\mu_4}{\mu_2}}, \quad (14)$$

$$w_0 = \frac{-2\mu_2^2\mu_4p + \mu_{2,2}\mu_2^2(p^2 - p) + 2\mu_0\mu_4^2}{2\mu_4^2}, \quad (15)$$

$$w_1 = \frac{\mu_2^2(\mu_4 + \mu_{2,2}(1 - p))}{2\mu_4^2}, \quad (16)$$

$$w_2 = \frac{\mu_2^2\mu_{2,2}}{4\mu_4^2}, \quad (17)$$

where $\mu_0 = \int 1 p d f_{\theta} d\theta$ is the zeroth central moment, $\mu_2 = \int \theta[i]^2 p d f_{\theta} d\theta$ the second central moment, $\mu_4 = \int \theta[i]^4 p d f_{\theta} d\theta$ the fourth central moment, and $\mu_{2,2} = \int \theta[i]^2 \theta[j \neq i]^2 p d f_{\theta} d\theta$ a bivariate central moment. Depending on the PDFs and their central moments, different PEM coefficients are derived. For instance, specifications of different PDFs and their corresponding PEM coefficients are given in Tables 1 and 2, respectively. Here, three different PDF types are considered: (i) normal distribution to represent the uncertainty range of well-known parameter variations; (ii) uniform distribution for cases where almost no information about the parameter variation is available; and (iii) triangle distribution to represent expert knowledge of parameter variations; i.e., the most likely parameter values and plausible upper and lower bounds. Within the individual distribution classes, any desired realization can be derived with a linear transformation step without loss of approximation accuracy, while non-symmetric or non-parametric distributions can be represented by iso-probabilistic but nonlinear transformation rules [42], which might introduce additional approximation errors.

Table 1. Statistical central moments of the standard normal (\mathcal{N}), uniform (\mathcal{U}), and triangle (\mathcal{T}) distribution.

Distribution	μ_0	μ_2	μ_4	$\mu_{2,2}$
$\theta[i] \sim \mathcal{N}(0, 1), \forall i = 1, \dots, p$	1	1	3	1
$\theta[i] \sim \mathcal{U}(-1, 1), \forall i = 1, \dots, p$	1	1/3	3/5	1/9
$\theta[i] \sim \mathcal{T}(-1, 1), \forall i = 1, \dots, p$	1	1/6	1/15	1/36

Table 2. PEM parameters according to the standard normal (\mathcal{N}), uniform (\mathcal{U}), and triangle (\mathcal{T}) distribution.

Distribution	ϑ	w_0	w_1	w_2
$\theta[i] \sim \mathcal{N}(0, 1), \forall i = 1, \dots, p$	$\sqrt{3}$	$1 + \frac{p^2-7p}{18}$	$\frac{4-p}{18}$	$\frac{1}{36}$
$\theta[i] \sim \mathcal{U}(-1, 1), \forall i = 1, \dots, p$	$\frac{3\sqrt{5}}{5}$	$1 + \frac{25p^2-295p}{1458}$	$\frac{160-25p}{1458}$	$\frac{25}{2916}$
$\theta[i] \sim \mathcal{T}(-1, 1), \forall i = 1, \dots, p$	$\frac{\sqrt{10}}{5}$	$1 + \frac{25p^2-145p}{288}$	$\frac{85-25p}{288}$	$\frac{25}{576}$

Once the deterministic sample points have been derived, the mean and the variance of the model simulations can be approximated as

$$E[y(t, \theta)] \approx \bar{y}(t, \theta) = w_0 y(t, \theta_0) + w_1 \sum_{k=1}^{2p} y(t, \theta_k) + w_2 \sum_{l=2p+1}^{2p^2} y(t, \theta_l), \quad (18)$$

$$\begin{aligned} \sigma^2(y(t, \theta)) \approx & w_0 (y(t, \theta_0) - \bar{y}(t, \theta))(y(t, \theta_0) - \bar{y}(t, \theta))^T + \\ & w_1 \sum_{k=1}^{2p} (y(t, \theta_k) - \bar{y}(t, \theta))(y(t, \theta_k) - \bar{y}(t, \theta))^T + \\ & w_2 \sum_{l=2p+1}^{2p^2} (y(t, \theta_l) - \bar{y}(t, \theta))(y(t, \theta_l) - \bar{y}(t, \theta))^T. \end{aligned} \quad (19)$$

Obviously, Equation (18) can directly be applied to approximate the averaged local sensitivity measure (Equation (7)), whereas the denominator of the Sobol indices (Equation (10)) is approximated with Equation (19). The calculation of the conditional variance terms of the numerator in Equation (10) is given in the following.

The overall set of generated sample points of the PEM provides nested subsets of sample points representing the conditional variance terms. According to the numerator of Equation (10), the expected value is taken over all but the i -th parameter; i.e., the ϑ -value of parameter i is set to zero. The variance, however, is taken exclusively for the i -th parameter; i.e., the corresponding ϑ entry is reset to its original distribution-dependent value. Only the weights, w_0 and w_1 , have to be adapted to a one-dimensional problem while leaving the overall PEM samples unchanged, i.e., there is no need for additional costly simulation runs. Please note the weight w_2 is set equal to zero; i.e., sample points generated via $GF(\vartheta, \theta)$ are not considered at this point. The described procedure is illustrated in Figure 2 assuming a three-dimensional parameter problem. In Figure 2a, the original sample set generated via Equation (11) is shown. Next, in Figure 2b, three different expected values are determined for $E[y(t)|\theta[1]]$ via the three differently colored sample subsets and different $\theta[1]$ values, respectively. When using these three derived expected values for $\theta[1]$, the conditional variance of $\theta[1]$ can be approximated (Equation (10)). The same procedure is repeated for parameter $\theta[2]$ (Figure 2c) and parameter $\theta[3]$ (Figure 2d) to derive the desired Sobol indices. Please note that the derived samples are not used to quantify the information content [16,46,47], but to quantify the uncertainty of the local sensitivities or to directly derive global parameter sensitivities. In [48,49], the performance of the PEM for calculating global parameter sensitivities and uncertainty analysis is discussed in more detail. Here, the PEM provides appropriate approximations at low computational costs compared to Monte Carlo simulations. In the

next step, from these sensitivity measures, the most MB-OED-relevant features have to be extracted by defining proper optimal design measures.

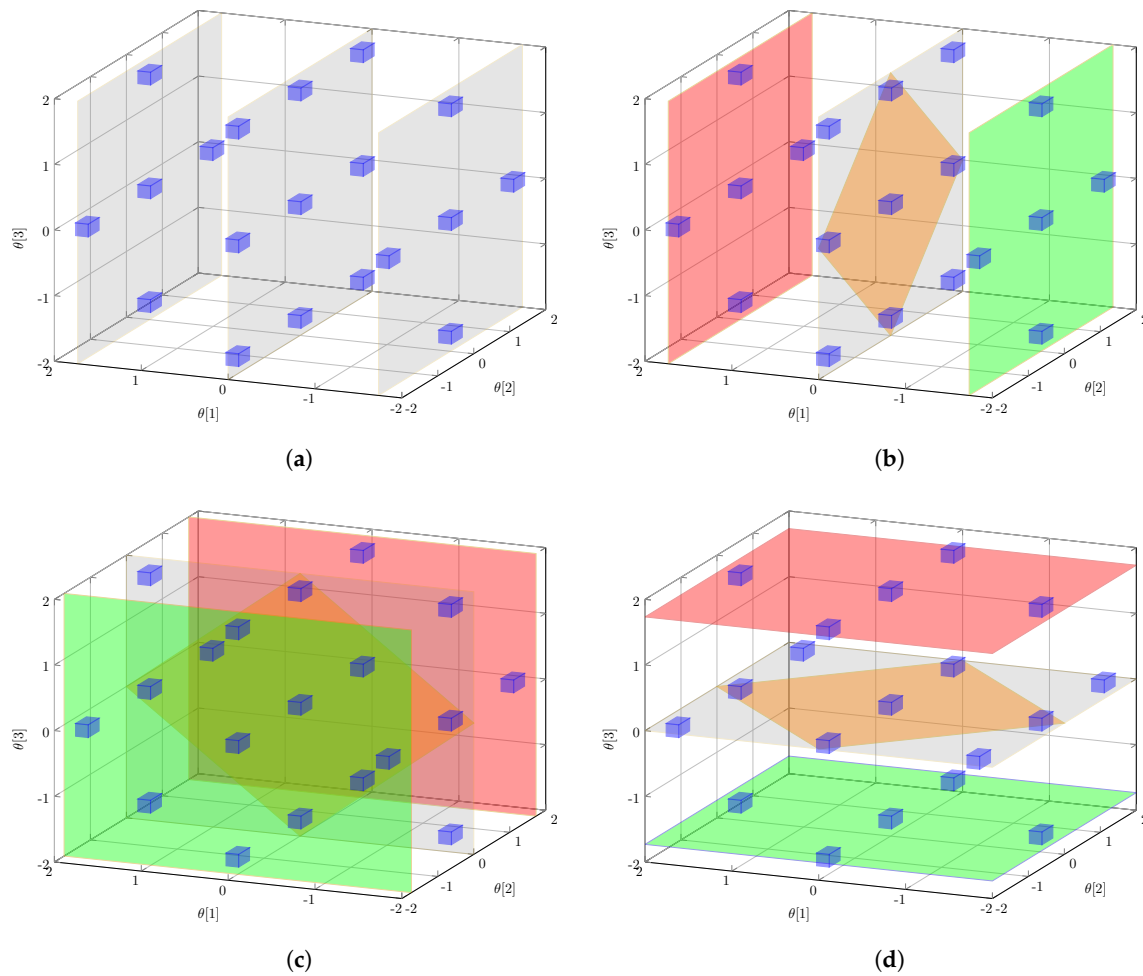


Figure 2. Parameter subset selection: Starting from an unconditioned parameter sample set (a), subsets are systematically selected in (b–d) to determine conditioned expectations and variances, respectively. (a) all sample points used to calculate the unconditioned, total uncertainty $\sigma^2(y(t))$; (b) sample subsets for $\theta[1]$, $\theta[1] + \delta$, and $\theta[1] - \delta$; (c) sample subsets for $\theta[2]$, $\theta[2] + \delta$, and $\theta[2] - \delta$; (d) sample subsets for $\theta[3]$, $\theta[3] + \delta$, and $\theta[3] - \delta$.

3. Optimal Design Measures

The actual optimization step of the MB-OED framework calls for a decision criterion that has to express the quality of the analyzed experimental configuration regarding parameter sensitivities. To this end, typically, representative scalar values are chosen for and evaluated with an optimization routine. In the literature, well-known criteria for local sensitivities exist and are summarized below followed by more general global design criteria.

3.1. Local Design Measures

The starting point for most local design criteria is an approximation of the parameter (co)variance matrix, which is defined as

$$C_\theta = E[(\theta - E[\theta])(\theta - E[\theta])^T]. \quad (20)$$

Based on the Cramér–Rao inequality [14] and local sensitivities (Equation (4)), the lower bound of the parameter (co)variance matrix assuming an unbiased estimator reads as

$$C_\theta \geq \text{FIM}^{-1}, \quad (21)$$

where the FIM is given by

$$\text{FIM} = \sum_{t_k}^K S^{\text{LSA}}(t_k)^T C_y^{-1}(t_k) S^{\text{LSA}}(t_k). \quad (22)$$

C_y represents the (co)variance matrix of the measurement data, which simplifies to a diagonal matrix assuming uncorrelated data samples. To formulate an optimization problem, a scalar function of the derived parameter (co)variance matrix C_θ is typically evaluated and minimized. Note that the inverse FIM is used as an approximation of the parameter covariance matrix in what follows, $C_\theta \approx \text{FIM}^{-1}$. A common class of these indicators is the so-called alphabetic family of design criteria [14,18] as given in Table 3. Here, λ_{\max} and λ_{\min} are the maximum and minimum eigenvalues of C_θ , respectively.

Table 3. Local design measures for MB-OED based on the parameter covariance matrix, where E_{LSA}^* refers to the modified E criterion.

Local Design Measures	Cost Functions
A_{LSA} —optimal design	$\Phi_A = \text{trace}(C_\theta)$
D_{LSA} —optimal design	$\Phi_D = \det(C_\theta)$
E_{LSA}^* —optimal design	$\Phi_E = \frac{\lambda_{\max}(C_\theta)}{\lambda_{\min}(C_\theta)}$

Which of these criteria leads to the best optimal experimental design is difficult to predict and depends on the analyzed problem at hand. Moreover, in many practical situations, the involved model parameters are highly correlated; for example, credible parameter estimates are challenging and sometimes impossible to derive for individual parameters. For that reason, dedicated anti-correlation criteria were proposed, aiming to reduce the parameter correlation while increasing the information content of the measurement data at the same time [50]. Here, linear parameter correlation is expressed as

$$\text{Corr}[i, j] = \frac{C_\theta[i, j]}{\sqrt{C_\theta[i, i]} \cdot \sqrt{C_\theta[j, j]}}, \quad (23)$$

with $i, j \in \{1, \dots, p\}$. The dominating correlation coefficients are selected directly or added as constraints to parameter variance measures [50], respectively. The basic notation reads as

$$\Phi_{\text{AC}}(\text{Corr}) = \text{Corr}[i, j]^2 + \text{Corr}[k, l]^2, \quad (24)$$

where $\text{Corr}[i, j]$ and $\text{Corr}[k, l]$ with $i, j, k, l \in \{1, \dots, p\}$ are the two most dominant correlation coefficients. As the (co)variance matrix C_θ is derived by the FIM (Equation (21)), these anti-correlation criteria depend on local parameter sensitivities (Equation (4)), too. As discussed in Section 2.1, local sensitivities may lead to crude approximation errors and sub-optimal experimental designs. In principle, the approximation error can be reduced by the multi-point averaging concept as proposed in Section 2.2 but still ignores nonlinear and multivariate effects as explained below.

While the anti-correlation strategy was successfully demonstrated for chemical and biochemical processes [50,51], it addresses linear parameter correlation solely, which, however, does not imply parameter independence [52]. Thus, nonlinear parameter correlations are neglected, but might be of relevance in MB-OED as well. In Figure 3, parameter dependencies are illustrated showing linear correlation (Figure 3a) and no correlation (Figure 3b) of two model parameters. In Figure 3c, however, the linear correlation coefficient (Equation (23)) is zero, but nonlinear interaction patterns are clearly visible. Ideally, the parameter estimate of one parameter should not impact the estimation outcome of other parameters. The same holds true for multivariate parameter interactions, i.e., the joint effect of more than two parameters that are also not considered by the linear anti-correlation measure

introduced in Equation (24). To address nonlinear parameter correlations and multivariate parameter interactions properly, global sensitivities and corresponding optimal design measures have to be analyzed instead.

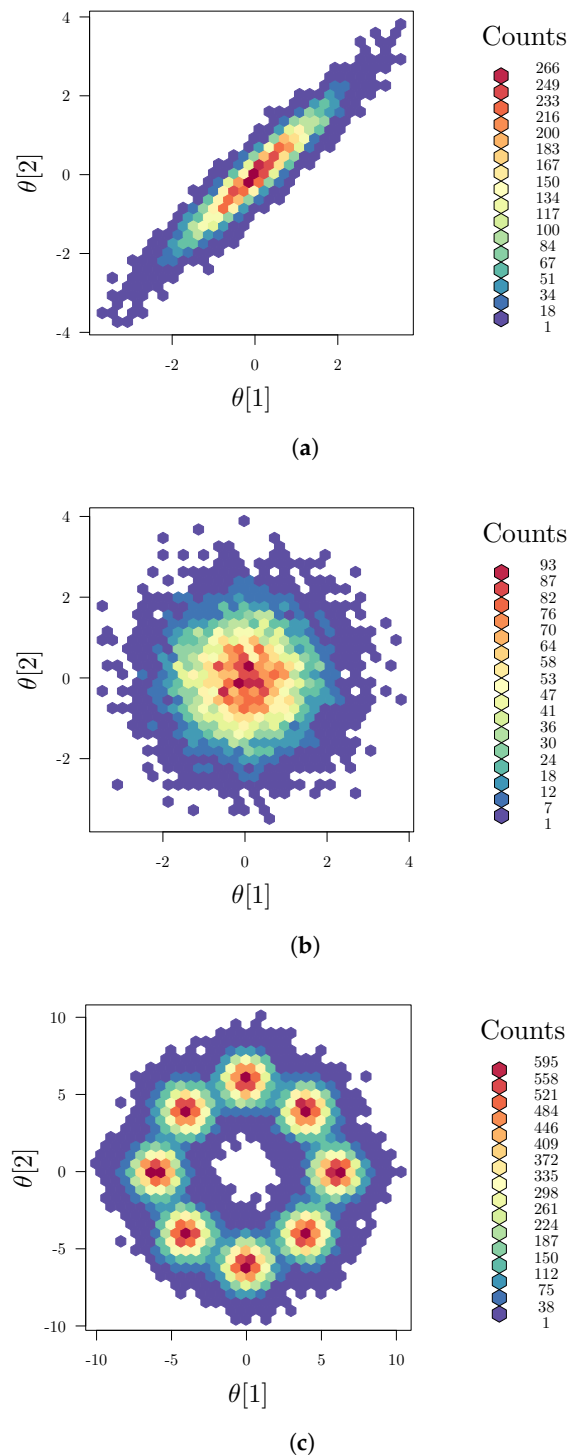


Figure 3. Illustrative example of parameter dependencies: (a) linear correlation with $\text{Corr}[1,2] = 0.95$; (b) no correlation with $\text{Corr}[1,2] = 0$; and (c) nonlinear correlation but zero linear correlation ($\text{Corr}[1,2] = 0$); here, eight standard Gaussian distributions are superimposed in a circular structure and properly rescaled providing a multi-modal density function. (a) linear correlated parameters; (b) linear uncorrelated parameters; (c) nonlinear correlation.

3.2. Global Design Measures

The additional insight given by global sensitivity measures (see Section 2.3) provides a new perspective on MB-OED [19,25,26,53]. As shown in Table 4, local design criteria (Table 3) are typically generalized to hold also for global sensitivities by substituting local sensitivities (Equation (4)) via global sensitivities (Equation (10)) [26,53].

Table 4. Generalized local design measures including global sensitivities.

GSA-Based Design Measures	Cost Functions
A_{GSA} —optimal design	$\Phi_{A_{\text{GSA}}} = \text{trace}([S^{\text{GSA}}]^T S^{\text{GSA}})^{-1}$
D_{GSA} —optimal design	$\Phi_{D_{\text{GSA}}} = \det([S^{\text{GSA}}]^T S^{\text{GSA}})^{-1}$
E_{GSA}^* —optimal design	$\Phi_{E_{\text{GSA}}^*} = \frac{\lambda_{\max}([S^{\text{GSA}}]^T S^{\text{GSA}})^{-1}}{\lambda_{\min}([S^{\text{GSA}}]^T S^{\text{GSA}})^{-1}}$

Thus, parameter uncertainties are taken into account naturally. Suboptimal results in MD-OED caused by misspecified reference points for local sensitivities (Equation (4)) can be avoided as much as possible by analyzing the entire PDF of the analyzed model parameters. Moreover, GSA provides additional insight into parameter dependencies, which can be included in the experimental design criteria as well. One attempt in this direction was made recently in [53], where so-called additive sensitivity indices are introduced as modified Sobol indices. Here, the key idea is to account for multivariate parameter interactions and dependencies that impact the overall quality of parameter estimates similar to individual parameter sensitivities. The usage of GSA to express generalized sensitivity matrices for MB-OED may pose some risk as well: (1) Not all GSA-based measures simplify to local measures for linear problems [25], which may lead to sub-optimal experimental designs and counter-intuitive outcomes; (2) including sensitivities of parameter combinations directly in a generalized GSA measure [53] might increase parameter dependencies; i.e., parameter correlations are amplified while limiting individual parameter contributions and, therefore, result in more challenging parameter identification problems; and (3) as the introduced GSA measures are normalized by the total variance factor, the calculated design may lead to an improved parameter sensitivity spectrum, but of lower sensitivities for individual parameters. Thus, the individual Sobol indices are optimized just by decreasing the $\sigma^2(t, y(t))$ term of Equation (10), meaning worse estimates for all model parameters.

To avoid the workaround of transferring global sensitivities into the local framework, the Shannon entropy as a general information measure [54] might be a helpful expression to calculate the desired features for global MB-OED. Intuitively, an optimal experimental design should fulfill the following features: (1) first, after running an experiment, the derived data have to ensure a balanced sensitivity of all parameters; i.e., all parameters should be practically identifiable. Thus, the corresponding Shannon entropy $\Phi_{\text{SE}_{\text{all}}}$ of the parameter sensitivities has to be at its maximum and the given criteria at its minimum (Table 5); i.e., there is a uniform distribution of the first-order Sobol indices. Note that $\overline{S^{\text{GSA}}}[i]$ refers to the averaged sensitivities covering all sample time points; (2) at a single time point t_k , however, it is desirable that only a very limited number of parameters contribute at the same time, i.e., to minimize (nonlinear) correlation effects. Thus, the Shannon entropy covering global sensitivities at a single time point $\Phi_{\text{SE}_{t_k}}$ has to be at its minimum; (3) to address multivariate parameter interactions and to minimize them, the gap between the sum of first-order Sobol indices and its theoretical maximum value of one is evaluated Φ_{PD} ; see Table 5. Here, a sum of first-order Sobol indices close to one corresponds to low parameter interactions, i.e., a well-posed parameter identification problem; and, (4) finally, the overall output variance Φ_{OU} has to be incorporated as well: the higher the better. A high output variance (Equation (19)) indicates a high parameter sensitivity, which is a prerequisite for credible parameter estimates. Please note that, for consistency, the related cost functions are given as a minimization problems.

Table 5. Global design measures for MB-OED based on Sobol indices and Shannon entropy.

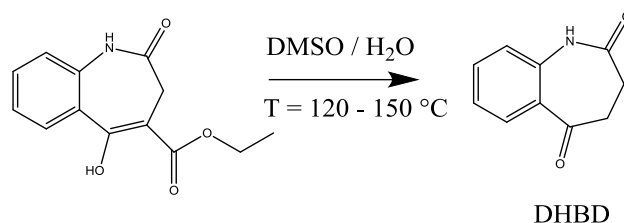
Global Design Measures	Cost Function
(1) Shannon entropy (entire time horizon)	$\Phi_{SE_{all}} = 1 / \sum_{i=1}^p \overline{S^{GSA}}[i] \ln(\overline{S^{GSA}}[i])$
(2) Shannon entropy (at single time point, t_k)	$\Phi_{SE_{t_k}} = \sum_{k=1}^K \sum_{i=1}^p S^{GSA}_{t_k}[i] \ln(S^{GSA}_{t_k}[i])$
(3) Parameter dependency	$\Phi_{PD} = \sum_{k=1}^K (1 - \sum_{i=1}^p S^{GSA}_{t_k}[i])$
(4) Overall output uncertainty	$\Phi_{OU} = 1 / \sum_{k=1}^K \sigma^2(y_{t_k})$

In the next step, MB-OED results based on the proposed global design measures are critically compared with the outcome of the traditional local sensitivity measures. To this end, various experimental design conditions are evaluated for two representative case studies in the field of pharmaceutical manufacturing.

4. Case Studies

4.1. Synthesis of an API-Scaffold (DHBD)

During the early stages of API process development, different properties of the unit operations involved are analyzed in order to characterize and to optimize the synthesis as well as the downstream route. Reaction rates are regarded as key descriptors of the synthesis progression as they depend on, temperature, and time. As a first case study for MB-OED, the homogeneous synthesis of 3,4-dihydro-1H-1-benzazepine-2,5-dione (DHBD) from the enolized 3-oxocarboxylic ester in wet dimethyl sulfoxide (DMSO) under neutral conditions and at elevated temperatures is presented; see Figure 4.

**Figure 4.** Synthesis of 3,4-dihydro-1H-1-benzazepine-2,5-dione (DHBD).

DHBD and its derivatives are pharmaceutically relevant scaffolds utilized for the synthesis of various protein kinase inhibitors and anticancer agents [55–60]. In traditional reaction optimization studies, isothermal syntheses at various temperatures are carried out one by one, and the syntheses are analyzed with offline high-performance liquid chromatography (HPLC) in order to determine reaction kinetics. However, HPLC measurements are tedious in sample preparation and need increased amounts of reactant, which might not be available in the very early stage of API development. Alternatively, we implement the fast-sampling and labor free in situ attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy in order to quantify the reactant concentration and, subsequently, to calculate reaction rate constants without the need for manual sampling. Combined with non-isothermal temperature profiles during the course of the reaction, temperature-dependent kinetic data, e.g., reaction rate constants at different temperatures and therefore Arrhenius parameters E_A and k_0 of the reaction under study, can be derived from a single experimental

run [61]. The quality of the derived reaction rate constants critically depends on the design of the applied non-isothermal temperature profile.

In what follows, we assume an Arrhenius rate expression:

$$k = k_0 \exp\left(-\frac{E_A}{RT}\right), \quad (25)$$

where k is the rate constant, T is the absolute temperature, R is the ideal gas constant, k_0 is the pre-exponential frequency factor, and E_A the activation energy. Because of the inherent correlation of the two Arrhenius parameters, k_0 and E_A , the parameterization of the Arrhenius equation (Equation (25)) is challenging. That is, independently of the applied experimental setup the correlation cannot be reduced for the Arrhenius rate [62,63] expression. At best, a parameter transformation might be applied to mitigate the correlation effect but changes the meaning of the identified parameters alike [64]. Thus, for this particular MB-OED problem, any anti-correlation criteria, which include the modified E^* -criterion as well, fail. Only the overall uncertainty of the parameter estimates can be reduced in principle. For this very reason, we first derive MB-OED results based on local sensitivities (Equation (4)) and the classical D-criterion (Section 3), which is expected to minimize the volume of the parameter confidence ellipsoid [14]:

$$T^*(t) = \arg \min_{T(t)} \Phi_D, \quad (26)$$

s.t.

$$\frac{dC_{DMSO}(t)}{dt} = -k(T(t))C_{DMSO}(t) \quad (27)$$

$$T_{lb} \leq T(t) \leq T_{ub}. \quad (28)$$

The optimal experimental design problem is described in Equations (26)–(28), where Equation (27) is the dynamic reaction model with the DMSO concentration $C_{DMSO}(t)$, and Equation (28) is the temperature constraint. A temperature profile divided into equidistant and constant subintervals is assumed. For all intervals, upper and lower temperature bounds are given as $T_{lb} = 100^\circ\text{C}$ and $T_{ub} = 150^\circ\text{C}$ while assuming continuous measurements of DHBD. Technically, the Matlab[®] (R2017a) optimizer `fmincon` is used to derive an optimal temperature profile; i.e., a profile that provides the most informative data and lowest parameter variations. The performance of the original and optimal temperature profile is validated with 2000 Monte Carlo simulations, where for each simulation and parameter identification, respectively, artificial experiment data are assumed with additive white noise. In Figure 5a, we show a reference temperature profile of five temperature steps of equal step-size, which was chosen by educated guessing. The reference profile might already be a good choice to identify the two Arrhenius parameters as it covers the whole temperature range without any obvious preferences. In Figure 5b, we see that the estimates are of finite variation, but as expected, they are strongly correlated. In Figure 5g,j, the individual parameter distributions represent the parameter uncertainties of E_A and k_0 from a different angle and represent their individual but finite variation. In the next step, the D-optimally designed temperature profile is derived numerically and shown in Figure 5b. Compared to the reference temperature profile, less dedicated temperature steps are visible over the experimental course. However, the range of the temperature values is lower; i.e., it starts at a higher temperature of 135°C and ends with the highest possible temperature of 150°C . Assuming the same measurement imperfections, the uncertainty in the identified Arrhenius parameters can be reduced by the optimized temperature profile; i.e., the uncertainty of the individual parameters is lower as indicated by the probability density functions in Figure 5h,k. The parameter correlation, however, as can be seen in Figure 5e, remains at the same level. As the derived optimal temperature profile might be difficult to realize due to the small temperature shifts resulting in operability and control issues, a simplified D^s-optimality temperature profile ignoring temperature shifts that are below

2 K variations (see Figure 5c) could be implemented with almost no performance loss; see Figure 5f,i,l for clarification. Similar lab-relevant constraints might be directly added to the underlying dynamic optimization problem [65] but are beyond the scope of this paper.

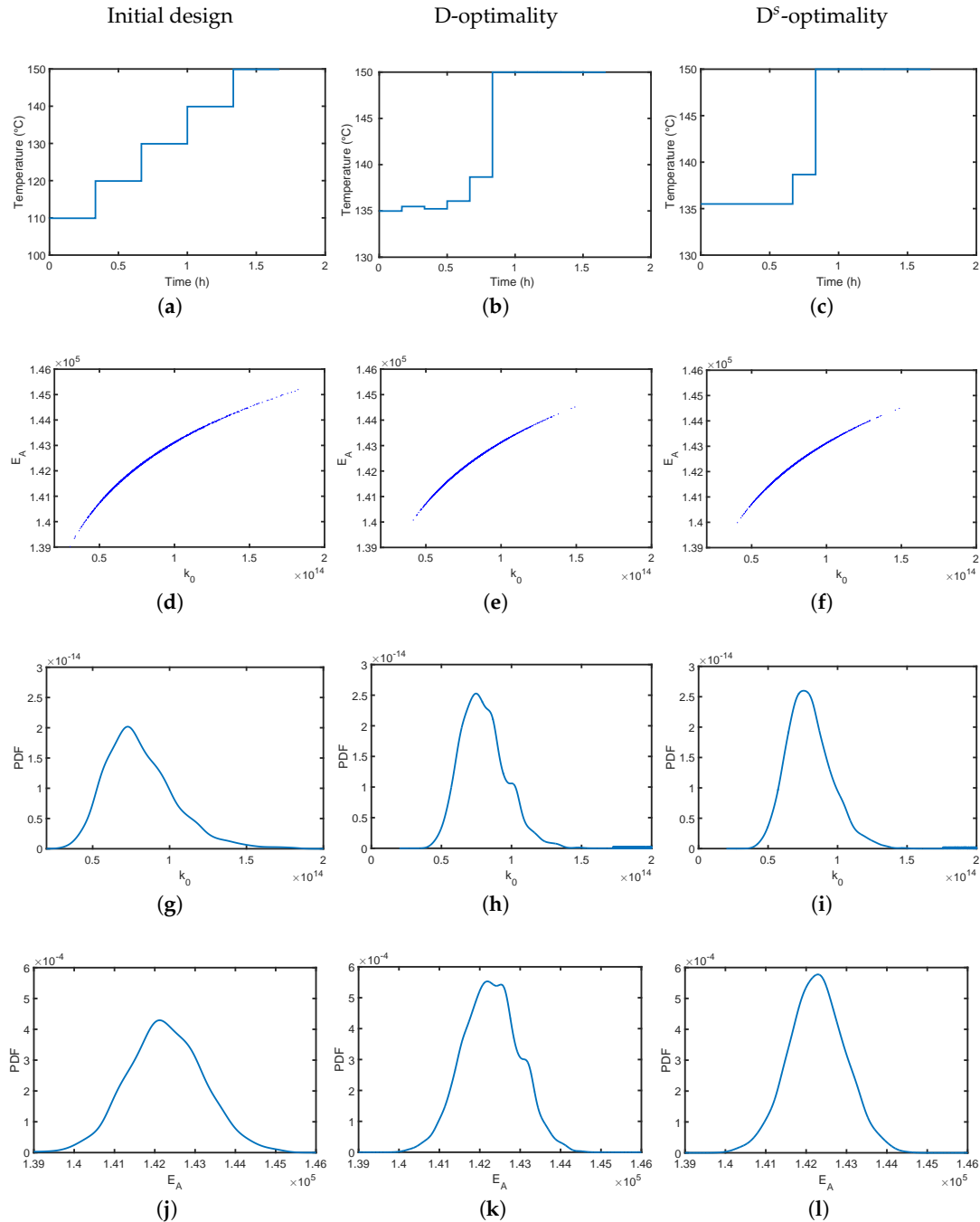


Figure 5. Performance of the initial setting and MB-OED. D^s-optimality means a simplified temperature profile in comparison to D-optimality. (a) is the initial; (b) the optimally designed; and (c) the optimal but simplified temperature profile; (d–f) are the scatter plots for the estimated parameters assuming simulated noisy data for 2000 MC simulations; (g–l) are the corresponding parameter probability distributions.

Thus far, only local sensitivities have been studied assuming the given reference Arrhenius parameters at which the local sensitivities have to be evaluated. The outcome of this local MB-OED

strategy, however, depends critically on the quality of the reference parameters at which the local sensitivities (Equation (4)) are derived. Any deviation of these parameters from their nominal values leads to a change in the local sensitivity values and the D-optimality. In Figure 6, the relative change in the optimized cost function is shown, which is defined as

$$\Delta\Phi_D^{\text{opt}}(k_0, E_A) = \frac{\Phi_D^{\text{opt}}(k_0, E_A) - \Phi_D^{\text{opt}}(k_0^{\text{ref}}, E_A^{\text{ref}})}{\Phi_D^{\text{opt}}(k_0^{\text{ref}}, E_A^{\text{ref}})}. \quad (29)$$

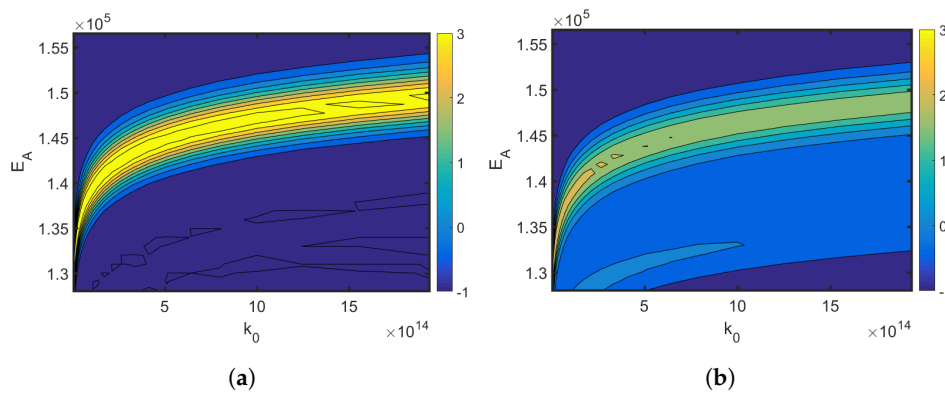


Figure 6. Effect of parameter variation. Areas close to zero mean no change in the optimized cost function (Equation (29)) when the Arrhenius parameters are changed. The classical D-optimality shown in (a) degrades drastically with parameter variations. The robust D-optimality shown in (b) is less affected by changes in the parameters. (a) D-optimality; (b) Robust D-optimality.

The results show that a misspecification of the applied reference Arrhenius parameters is likely to result in sub-optimal temperature profiles, less informative measurement data, and higher parameter uncertainties. Alternatively, when the multi-point averaging approach is used (Equation (8)), the calculated design is more robust against reference parameter variations; see Figure 6b. The classical D-optimality shown in Figure 6a degrades drastically with reference variations. The robust D-optimality, however, shown in Figure 6b is less affected by changing the reference parameters. As the reference parameters are typically unknown but are needed to calculate local parameter sensitivities, a robust MB-OED strategy is expected to provide more valuable MB-OED results. Moreover, global parameter sensitivities are evaluated and used for a more credible D-optimality measure (Table 4). The resulting parameter errors for the classical D-optimality, its multi-point averaging realization, and the global sensitivity-based D-design are analyzed. In Figure 7, the multi-point averaging approach and the GSA-based MB-OED lead to more precise parameter estimates in comparison to the classical D-optimality as indicated by the density functions. Here, the GSA-based design results in the most precise estimates; i.e., the probability density functions of Arrhenius parameter have their highest peak close to the true values. Moreover, the multi-point averaging approach and the GSA-based design seem to be less corrupted by the misleading second local minima of the parameter identification problem, which is indicated by the second peak of the probability density functions in Figure 7a,b. Please note, because the proposed PEM sampling strategy is used, only nine sample points for each iteration of the optimizations step are needed to calculate the multi-point averaging or GSA measures in the case of the two Arrhenius parameters.

The applied sensitivity measure has a strong impact on the MB-OED results for the Arrhenius parameters. The classical MB-OED based on local sensitivities is error-prone and is expected to provide sub-optimal experimental designs. For novel APIs, which are available only in very small quantities, each individual experimental run counts. Therefore, MB-OED following a multi-point averaging approach or GSA principles seems to be preferable. Combined with the proposed non-isothermal temperature profile strategy, the optimized temperature profiles ensure the best use of the API-scaffold

DHBD and the most precise estimates of the kinetic rate parameters, k_0 and E_A . In the next step, an MB-OED study of a more complex biochemical synthesis problem is presented where the focus is also on the effect of global design measures.

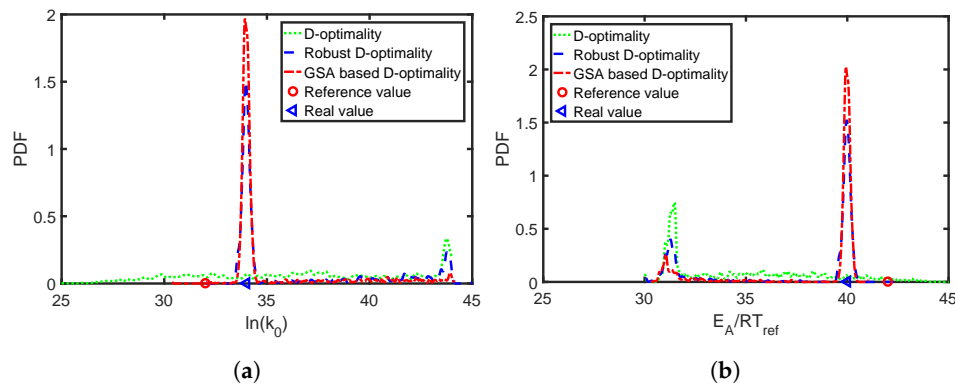


Figure 7. The resulting parameter distributions of the two Arrhenius parameters for the discussed MB-OED strategies. The red circle is the reference value used for MB-OED while the blue triangle is the true parameter value. (a) Arrhenius parameter k_0 ; (b) Arrhenius parameter E_A .

4.2. A Fed-Batch Bioreactor

In the second test case study, a fed-batch bioreactor is analyzed. A lumped version of a generic biomass-substrate model reads as follows:

$$\frac{dC_X}{dt} = \mu \cdot C_X - \frac{U}{V} \cdot C_X, \quad (30)$$

$$\frac{dC_S}{dt} = -\sigma \cdot C_X + \frac{U}{V} \cdot (C_{S,in} - C_S), \quad (31)$$

$$\frac{dV}{dt} = U, \quad (32)$$

where C_X is the biomass concentration, C_S is the substrate concentration, V is the liquid volume, and U is the inlet flow rate. The specific growth rate μ and the specific consumption rate σ follow Monod expressions

$$\mu = \mu_{max} \cdot \frac{C_S}{C_S + K_S}, \quad (33)$$

$$\sigma = \frac{\mu}{Y_{X/S} + m}. \quad (34)$$

The nominal parameter values in Equations (30)–(34) and the initial conditions for the dynamic model are listed in Table 6.

Table 6. Parameters as in [66] and initial conditions for the fed-batch bioreactor.

Symbol	Parameter	Unit	Nominal Value
C_{X0}	initial concentration of biomass	g L^{-1}	0.25
C_{S0}	initial concentration of substrate	g L^{-1}	3
V_0	initial volume of the liquid phase	L	7
$C_{S,in}$	substrate concentration in the feed	g L^{-1}	50
U	volumetric feed rate	L h^{-1}	0–1
μ_{max}	maximum specific growth rate	h^{-1}	0.421
K_S	half velocity constant	g L^{-1}	0.439
$Y_{X/S}$	yield coefficient of biomass over substrate	-	0.777
m	maintenance factor	h^{-1}	0

For the experimental setups, the following constraints are considered. The duration of the reaction is set to 15 h. The biomass and the substrate concentration are measurable with a sample rate of 0.75 h, which results in a total set of 40 measurement samples. The maximum specific growth rate μ_{max} , the half velocity constant K_S , and the yield coefficient $Y_{X/S}$ should be estimated by minimizing the difference between the measurements and the simulation results. In this simulation study, artificial measurement data [67] are used with additive measurement noise of $\mathcal{N}(0, 6.25 \times 10^{-4})$ and $\mathcal{N}(0, 1 \times 10^{-3})$ for C_x and C_s , respectively. The MB-OED strategy aims at reducing the uncertainty of the parameters by optimizing the feeding policy of the fed-batch bioreactor. Thus, the inlet profile U is parameterized by 20 constant segments of equal size, which are optimized to provide the most informative experimental data. The Monod kinetic parameters, μ_{max} , K_S and $Y_{X/S}$, are treated as uncertain. Their variations are expressed by uniform PDFs of different ranges, i.e., $\mu_{max} \sim U(0.2, 0.8)$, $K_S \sim U(0.2, 0.8)$, and $Y_{X/S} \sim U(0.4, 1.5)$, which might have been derived experimentally with a parameter identification procedure. In addition to μ_{max} , K_S , and $Y_{X/S}$, the model parameters are assumed to be given by the literature without any uncertainty. Thus, please note that only 19 sample points for each iteration of the optimization step are needed to calculate the multi-point averaging or GSA measures when using the proposed PEM sampling strategy.

First, we applied the E^* -design to minimize imbalanced parameter sensitivities and uncertainty. Similar to the previous case study, we compare a default substrate inlet profile (Figure 8c) with the outcome of the classical E^* -design in Figure 8d–f. The resulting parameter scatter plots based on the initial profile (Figure 9a–c) show bimodal behavior indicating some severe nonlinearity and a non-convex optimization problem. For the classical E^* -design, the resulting parameter variations of all three parameters (Figure 9d–f) could be only slightly improved while the parameter correlation for all parameter combinations is clearly visible. The multi-point averaging approach of the E^* -design, in turn, seems to be more appropriate. Applying a low initial substrate concentration (Figure 8g) combined with an optimized impulse-like feeding rate (Figure 8i), the quality of the parameter estimates can be improved as illustrated in Figure 9g–i. In particular, the parameter correlations between K_S and $Y_{X/S}$, and μ_{max} and K_S are reduced. This effect can even be improved by implementing an MB-OED strategy based on GSA. A low initial substrate concentration (Figure 8j) but a more complex feeding rate (Figure 8l) results in more compact parameter scatter plots (Figure 9j–l); i.e., fewer parameter uncertainties and parameter correlations. Finally, in the Shannon 1 study, we also analyze the performance of a Shannon-entropy-based (Table 5) multi-objective design using $w = [1, 1, 0, 1]^T$ as a weighting vector for

$$\Phi_{SE} = w^T [\Phi_{SE_{all}}, \Phi_{SE_{t_k}}, \Phi_{PD}, \Phi_{OU}]. \quad (35)$$

Please note that the third element $w[3]$ was set to zero as the sum of the first-order Sobol indices ($\sum_{i=1}^p S_{GSA_{t_k}}^{SA}[i]$) was always greater or equal to 0.9; i.e., there is no strong multivariate dependency of the analyzed parameters.

Similar to the E^* -design, the derived initial substrate concentration is high; see Figure 8m. For the feeding rate, there are two distinct feeding periods as illustrated in Figure 8o. The resulting parameter estimates, which are summarized in Figure 9m–o, show no significant improvement in the parameter estimates in comparison to the unoptimized initial experimental setting or the E^* -design. The main reason here is the tedious tuning of the multi-objective function, i.e., providing an optimal weighting vector w . For instance, when using $w = [1000, 1, 0, 0.1]^T$ in the Shannon 2 study, the quality of the parameter estimates improves significantly as illustrated in Figure 10. In comparison to the previous MB-OED results, the Shannon-entropy-based design in combination with a proper weight vector w leads to improved parameter estimates, i.e., lower parameter variations and correlations as shown in Figure 10d–f. A low initial substrate concentration (Figure 10a) followed by a gradual increase in the feeding rate (Figure 10c) provide very informative data and precise parameter estimates, respectively. This conclusion can be validated by analyzing Figure 11. The relative parameter errors $|\hat{\theta} - \theta|/\theta$ clearly indicate the second Shannon-entropy-based design as the most suitable one for the

parameter identification problem. Please note parameters θ were sampled from their assumed uniform density functions.

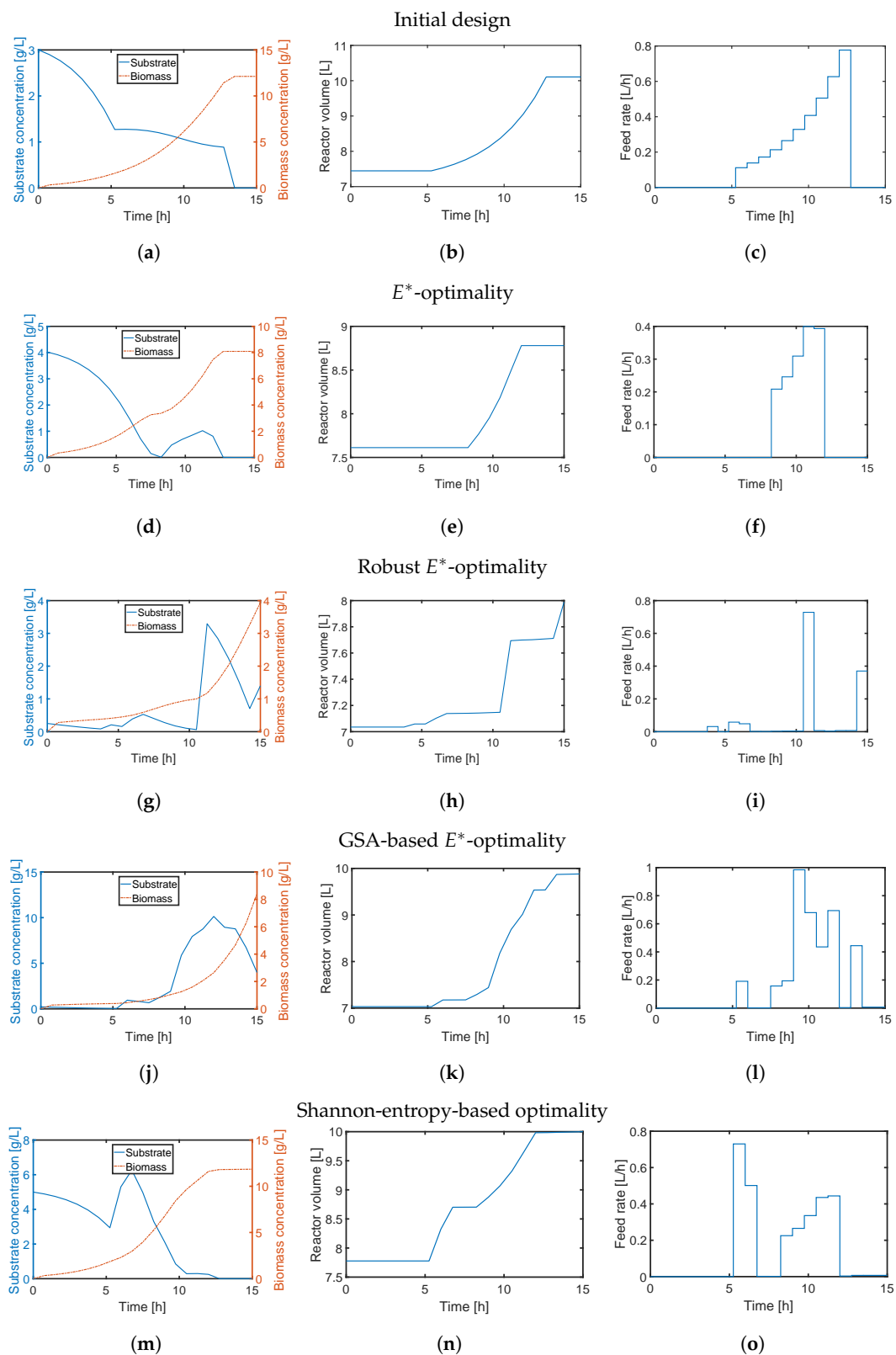


Figure 8. MB-OED results showing the evolution of biomass and substrate concentration (**left**), and volume of the reactor (**middle**) based on the optimized feed rate (**right**).

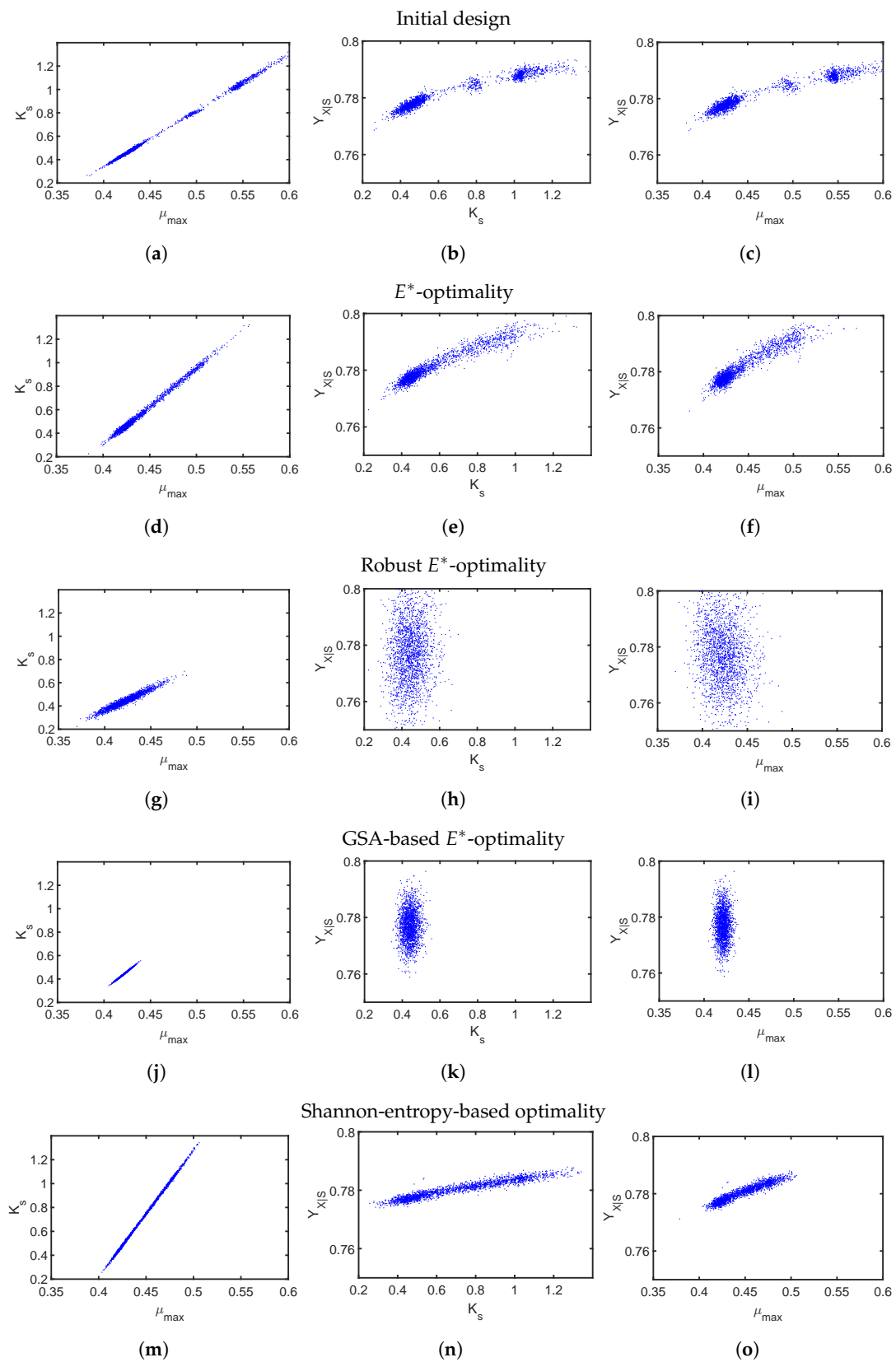


Figure 9. Performance of the experiments from initial design (a–c), E^* -optimality (d–f), robust E^* -optimality (g–i), GSA-based E^* -optimality (j–l), and Shannon-entropy-based design (m–o), where scatter plots are used to present the confidence region for the estimated parameters, which result from 2000 MC simulations.

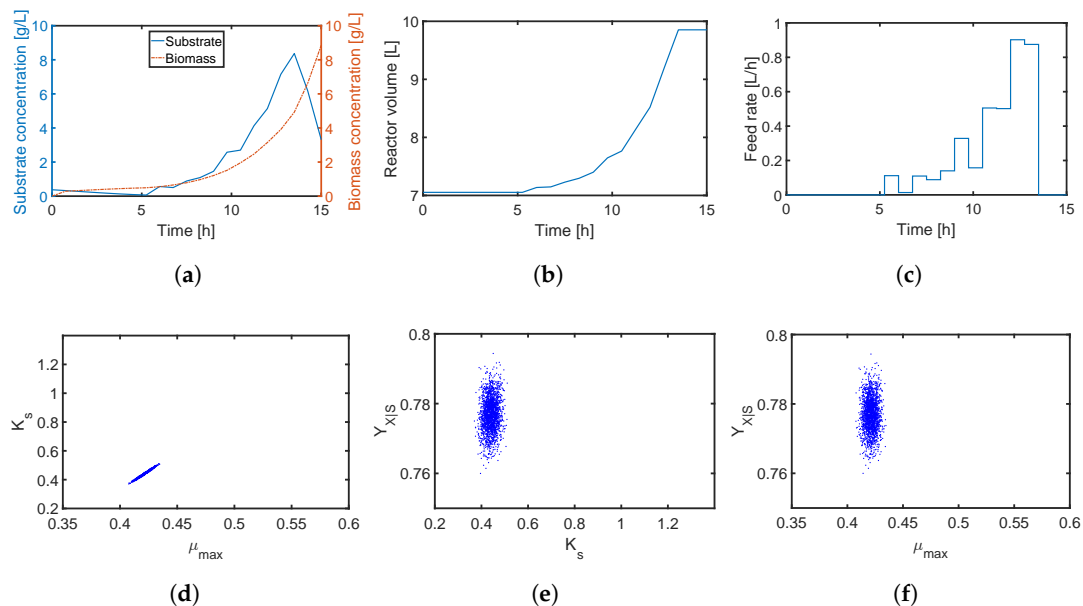


Figure 10. Evolution profiles for three states ((a) biomass concentration, substrate concentration; and (b) volume of the reactor) and (c) the feed rate for the adapted Shannon-entropy-based design with $w = [1000, 0.1, 0, 1]^T$. In (d–f), the resulting parameter estimates are shown.

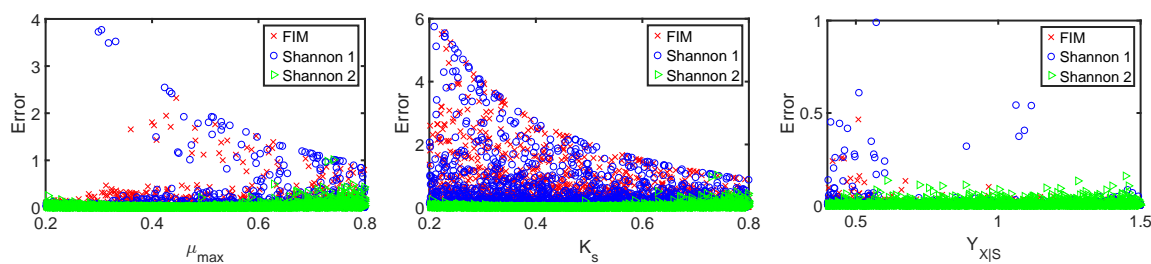


Figure 11. Error of the parameter estimates of μ_{max} , K_S and $Y_{X/S}$. 2000 MC simulations are shown for the classical E^* -design and Shannon-entropy-based design with $w = [1, 1, 0, 1]$, and $w = [1000, 1, 0, 0.1]$.

5. Conclusions

Without a doubt, model-based and computer-aided concepts are valuable tools for translating promising lab findings into efficient processes. Biased model-based results, however, due to model misspecification and model parameter uncertainties, are equally probable. This is particularly true when measurement data are limited in quantity and quality. Thus, in our work, we have successfully demonstrated the benefit of robustification and GSA concepts for MB-OED to gain the most informative data and to improve parameter estimates, respectively. The classical MB-OED, which is based on local sensitivities, is error-prone and is expected to provide sub-optimal experimental designs. For novel APIs, which are available only in very small quantities, each individual experimental run counts. Therefore, MB-OED following a multi-point averaging approach or GSA principles seem to be preferable as demonstrated by the DHBD synthesis problem. Moreover, Shannon-entropy-based design measures, which can account for global parameter sensitivities easily, provide a new angle in MB-OED. Assuming a proper weighting factor, it provides low parameter uncertainties and correlations as illustrated by the biochemical synthesis study.

To ensure the practical relevance, an efficient implementation of the proposed ideas was of particular interest in this study. The point estimate method framework seems to be an attractive alternative compared to state-of-the-art Monte Carlo simulations. On the one hand, it significantly

reduces the computational load for advanced sensitivity analyses. On the other hand, its ease in implementation ensures a straightforward adaptation for various problems in MB-OED.

Nevertheless, there are still some open issues and potential research questions that have to be addressed in ongoing research. For instance, there is no systematic analysis under which conditions global sensitivity or robustification concepts perform best and, therefore, need to be explored in greater detail. This could lead to best-practice rules and informed decisions in MB-OED for nonlinear and complex models. An intensive study of tailored design metrics reflecting global sensitivity outcomes may also help to gain improvements in the model building process. In particular, multi-objective design concepts have to be explored and validated in the future. Finally, the proposed ideas have to be systematically extended to incorporate model misspecification as well, i.e., uncertainties related to the model structure itself in addition to the model parameters.

Acknowledgments: We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the Technische Universität Braunschweig. Funding of the “Promotionsprogramm μ -Props” for Xiangzhong Xie and Moritz Rehbein by MWK Niedersachsen, Germany is gratefully acknowledged. Moreover, the anonymous reviewers deserve special thanks for their valuable input.

Author Contributions: R.S., X.X., and M.R. collected information and wrote the manuscript. R.S. conceived the research. X.X. performed the simulations. U.K. and S.S. provided feedback to the content and participated in writing the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Franz, A.; Song, H.S.; Ramkrishna, D.; Kienle, A. Experimental and theoretical analysis of poly(β -hydroxybutyrate) formation and consumption in *Ralstonia eutropha*. *Biochem. Eng. J.* **2011**, *55*, 49–58.
2. Grimard, J.; Dewasme, L.; Wouwer, A.V. A review of dynamic models of hot-melt extrusion. *Processes* **2016**, *4*, 19, doi:10.3390/pr4020019.
3. Bück, A.; Neugebauer, C.; Meyer, K.; Palis, S.; Diez, E.; Kienle, A.; Heinrich, S.; Tsotsas, E. Influence of operation parameters on process stability in continuous fluidised bed layering with external product classification. *Powder Technol.* **2016**, *300*, 37–45.
4. Logist, F.; Erdeghem, P.V.; Smets, I.Y.; Impe, J.F.V. Multiple-objective optimisation of a jacketed tubular reactor. In Proceedings of the 2007 European Control Conference (ECC), Kos, Greece, 2–5 July 2007; pp. 963–970.
5. Kunde, C.; Michaels, D.; Micovic, J.; Lutze, P.; Górak, A.; Kienle, A. Deterministic global optimization in conceptual process design of distillation and melt crystallization. *Chem. Eng. Process. Process Intensif.* **2016**, *99*, 132–142.
6. Kaiser, N.M.; Flassig, R.J.; Sundmacher, K. Probabilistic reactor design in the framework of elementary process functions. *Comput. Chem. Eng.* **2016**, *94*, 45–59.
7. Föste, H.; Schöler, M.; Majschak, J.P.; Augustin, W.; Scholl, S. Modeling and validation of the mechanism of pulsed flow cleaning. *Heat Transf. Eng.* **2013**, *34*, 753–760.
8. Schenkendorf, R. Supporting the shift towards continuous pharmaceutical manufacturing by condition monitoring. In Proceedings of the 2016 IEEE 3rd Conference on Control and Fault-Tolerant Systems (SysTol), Barcelona, Spain, 7–9 September 2016.
9. Jelemenski, M.; Klauo, M.; Paulen, R.; Lauwers, J.; Logist, F.; Impe, J.V.; Fikar, M. Time-optimal control and parameter estimation of diafiltration processes in the presence of membrane fouling. *IFAC-PapersOnLine* **2016**, *49*, 242–247.
10. Rogers, A.J.; Inamdar, C.; Ierapetritou, M.G. An integrated approach to simulation of pharmaceutical processes for solid drug manufacture. *Ind. Eng. Chem. Res.* **2014**, *53*, 5128–5147.
11. Rantanen, J.; Khinast, J. The future of pharmaceutical manufacturing sciences. *J. Pharm. Sci.* **2015**, *104*, 3612–3638.
12. Ljung, L. *System Identification: Theory for the User*; Prentice-Hall: Upper Saddle River, NJ, USA, 1987.
13. Balsa-Canto, E.; Alonso, A.A.; Banga, J.R. An iterative identification procedure for dynamic modeling of biochemical networks. *BMC Syst. Biol.* **2010**, *4*, 11, doi:10.1186/1752-0509-4-11.

14. Walter, E.E.; Pronzato, L. *Identification of Parametric Models from Experimental Data*; Springer: New York, NY, USA, 1997; p. 413.
15. Rodríguez-Aragón, L.J.; López-Fidalgo, J. Optimal designs for the Arrhenius equation. *Chemom. Intell. Lab. Syst.* **2005**, *77*, 131–138.
16. Schenkendorf, R.; Kremling, A.; Mangold, M. Optimal experimental design with the sigma point method. *IET Syst. Biol.* **2009**, *3*, 10–23.
17. Galvanin, F.; Marchesini, R.; Barolo, M.; Bezzo, F.; Fidaleo, M. Optimal design of experiments for parameter identification in electrodialysis models. *Chem. Eng. Res. Des.* **2016**, *105*, 107–119.
18. Kiefer, J. Optimum experimental designs. *J. R. Stat. Soc. Ser. B* **1959**, *21*, 272–319.
19. Martínez, E.C.; Cristaldi, M.D.; Grau, R.J. Design of Dynamic Experiments in Modeling for Optimization of Batch Processes. *Ind. Eng. Chem. Res.* **2009**, *48*, 3453–3465.
20. Sinkoe, A.; Hahn, J. Optimal experimental design for parameter estimation of an IL-6 signaling model. *Processes* **2017**, *5*, 49, doi:10.3390/pr5030049.
21. Manesso, E.; Sridharan, S.; Gunawan, R. Multi-objective optimization of experiments using curvature and Fisher information matrix. *Processes* **2017**, *5*, 63, doi:10.3390/pr5040063.
22. Chaloner, K.; Verdinelli, I. Bayesian experimental design: A review. *Stat. Sci.* **1995**, *10*, 273–304.
23. Gotwalt, C.M.; Jones, B.A.; Steinberg, D.M. Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics* **2009**, *51*, 88–95.
24. Overstall, A.M.; Woods, D.C. Bayesian design of experiments using approximate coordinate exchange. *Technometrics* **2017**, *59*, 458–470.
25. Chu, Y.; Hahn, J. Necessary condition for applying experimental design criteria to global sensitivity analysis results. *Comput. Chem. Eng.* **2013**, *48*, 280–292.
26. Rodríguez-Fernandez, M.; Kucherenko, S.; Pantelides, C.; Shah, N. Optimal experimental design based on global sensitivity analysis. *Comput. Aided Chem. Eng.* **2007**, *24*, 63–68.
27. Bockstal, P.J.V.; Mortier, S.; Corver, J.; Nopens, I.; Gernaey, K.V.; Beer, T.D. Global sensitivity analysis as good modelling practices tool for the identification of the most influential process parameters of the primary drying step during freeze-drying. *Eur. J. Pharm. Biopharm.* **2018**, *123*, 108–116.
28. Scire, J., Jr.; Dryer, F.; Yetter, R. Comparison of global and local sensitivity techniques for rate constants determined using complex reaction mechanisms. *Int. J. Chem. Kinet.* **2001**, *33*, 784–802.
29. Saltelli, A.; Ratto, M.; Tarantola, S.; Campolongo, F. Sensitivity analysis for chemical Models. *Chem. Rev.* **2005**, *105*, 2811–2828.
30. Saltelli, A.; Aleksankina, K.; Becker, W.; Fennell, P.; Ferretti, F.; Holst, N.; Sushan, L.; Wu, Q. Why so many published sensitivity analyses are false. A systematic review of sensitivity analysis practices. *arXiv* **2017**, arXiv:1711.11359.
31. Rao, M.M.; Swift, R.J. *Probability Theory with Applications*; Springer: New York, NY, USA, 2006.
32. Turanyi, T. Sensitivity Analysis of Complex Kinetic Systems. Tools and Applications. *J. Math. Chem.* **1990**, *5*, 203–248.
33. Bauer, I.; Bock, H.G.; Körkel, S.; Schlöder, J.P. Numerical methods for optimum experimental design in DAE systems. *J. Comput. Appl. Math.* **2010**, *120*, 1–25.
34. Ten Broeke, G.; van Voorn, G.; Kooi, B.; Molenaar, J. Detecting tipping points in ecological models with sensitivity analysis. *Math. Model. Nat. Phenom.* **2016**, *11*, 47–72.
35. Kucherenko, S.; Rodríguez-Fernandez, M.; Pantelides, C.; Shah, N. Monte Carlo evaluation of derivative-based global sensitivity measures. *Reliab. Eng. Syst. Saf.* **2009**, *94*, 1135–1148.
36. Rakovec, O.; Hill, M.C.; Clark, M.P.; Weerts, A.H.; Teuling, A.J.; Uijlenhoet, R. Distributed evaluation of local sensitivity analysis (DELSA), with application to hydrologic models. *Water Resour. Res.* **2014**, *50*, 409–426.
37. Zádor, J.; Zsély, I.; Turányi, T. Local and global uncertainty analysis of complex chemical kinetic systems. *Reliab. Eng. Syst. Saf.* **2006**, *91*, 1232–1240.
38. Iooss, B.; Lemaître, P. A review on global sensitivity analysis methods. *arXiv* **2014**, *23*, arXiv:1404.2405.
39. Sobol, I. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **2001**, *55*, 271–280.
40. Buzzard, G.T. Global sensitivity analysis using sparse grid interpolation and polynomial chaos. *Reliab. Eng. Syst. Saf.* **2012**, *107*, 82–89.
41. Caflisch, R.E. Monte Carlo and quasi-Monte Carlo methods. *Acta Numer.* **1998**, *7*, 1–49.

42. Sudret, B. *Uncertainty Propagation and Sensitivity Analysis in Mechanical Models Contributions to Structural Reliability and Stochastic Spectral Methods*; Habilitation a diriger des recherches; Université Blaise Pascal: Clermont-Ferrand, France, 2007.
43. Tyler, G.W. Numerical integration of functions of several variables. *Can. J. Math.* **1953**, *5*, 393–412.
44. Lerner, U.N. Hybrid Bayesian Networks for Reasoning about Complex Systems. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 2002.
45. Evans, D.H. An application of numerical integration techniques to statistical tolerancing. *Technometrics* **1967**, *9*, 441–456.
46. Joshi, M.; Seidel-Morgenstern, A.; Kremling, A. Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metab. Eng.* **2006**, *8*, 447–455.
47. Heine, T.; Kawohl, M.; King, R. Derivative-free optimal experimental design. *Chem. Eng. Sci.* **2008**, *63*, 4873–4880.
48. Schenkendorf, R.; Mangold, M. Qualitative and quantitative optimal experimental design for parameter identification of a MAP kinase model. *IFAC Proc. Vol.* **2011**, *44*, 11666–11671.
49. Schenkendorf, R.; Mangold, M. Online model selection approach based on unscented Kalman filtering. *J. Process Control* **2013**, *23*, 44–57.
50. Franceschini, G.; Macchietto, S. Novel anticorrelation criteria for design of experiments: Algorithm and application. *AIChE J.* **2008**, *54*, 3221–3238.
51. Ohs, R.; Wendlandt, J.; Spiess, A.C. How graphical analysis helps interpreting optimal experimental designs for nonlinear enzyme kinetic models. *AIChE J.* **2017**, *63*, 4870–4880.
52. Embrechts, P.; Mcneil, A.; Straumann, D. Correlation and dependence in risk management: Properties and pitfalls. In *RISK Management: Value at Risk and Beyond*; Cambridge University Press: Cambridge, UK, 2002; pp. 176–223.
53. Kucerová, A.; Šýkora, J.; Janouchová, E.; Jarušková, D.; Chleboun, J. Acceleration of robust experiment design using Sobol indices and polynomial chaos expansion. In Proceedings of the 7th International Workshop on Reliable Engineering Computing (REC), Bochum, Germany, 15–17 June 2016; pp. 411–426.
54. Lindley, D.V. On a measure of the information provided by an experiment. *Ann. Math. Stat.* **1956**, *27*, 986–1005.
55. Becker, A.; Kohfeld, S.; Lader, A.; Preu, L.; Pies, T.; Wieking, K.; Ferandin, Y.; Knockaert, M.; Meijer, L.; Kunick, C. Development of 5-benzylpaullones and paullone-9-carboxylic acid alkyl esters as selective inhibitors of mitochondrial malate dehydrogenase (mMDH). *Eur. J. Med. Chem.* **2010**, *45*, 335–342.
56. Egert-Schmidt, A.M.; Dreher, J.; Dunkel, U.; Kohfeld, S.; Preu, L.; Weber, H.; Ehlert, J.E.; Mutschler, B.; Totzke, F.; Schächtele, C.; et al. Identification of 2-Anilino-9-methoxy-5,7-dihydro-6H-pyrimido[5,4-d][1]benzazepin-6-ones as Dual PLK1/VEGF-R2 Kinase Inhibitor Chemotypes by Structure-Based Lead Generation. *J. Med. Chem.* **2010**, *53*, 2433–2442.
57. Falke, H.; Chaikuad, A.; Becker, A.; Loaëc, N.; Lozach, O.; Jhaisha, S.A.; Becker, W.; Jones, P.G.; Preu, L.; Baumann, K.; et al. 10-Iodo-11H-indolo[3,2-c]quinoline-6-carboxylic Acids Are Selective Inhibitors of DYRK1A. *J. Med. Chem.* **2015**, *58*, 3131–3143.
58. Kunick, C. Fused azepinones with antitumor activity. *Curr. Pharm. Des.* **1999**, *5*, 181–194.
59. Tolle, N.; Kunick, C. Paullones as Inhibitors of Protein Kinases. *Curr. Top. Med. Chem.* **2011**, *11*, 1320–1332.
60. Kunick, C.; Bleeker, C.; Prühs, C.; Totzke, F.; Schächtele, C.; Kubbutat, M.H.; Link, A. Matrix compare analysis discriminates subtle structural differences in a family of novel antiproliferative agents, diaryl-3-hydroxy-2,3,3a,10a-tetrahydrobenzo[b]cyclopenta[e]azepine-4,10(1H,5H)-diones. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2148–2153.
61. Rehbein, M.C.; Husmann, S.; Lechner, C.; Kunick, C.; Scholl, S. Fast and calibration free determination of first order reaction kinetics in API synthesis using in situ ATR-FTIR. *Eur. J. Pharm. Biopharm.* **2017**, in press.
62. Varga, L.; Szabó, B.; Zsély, I.G.; Zempléni, A.; Turányi, T. Numerical investigation of the uncertainty of Arrhenius parameters. *J. Math. Chem.* **2011**, *49*, 1798–1809.
63. Nagy, T.; Turányi, T. Uncertainty of Arrhenius parameters. *Int. J. Chem. Kinet.* **2011**, *43*, 359–378.
64. Schwaab, M.; Lemos, L.P.; Pinto, J.C. Optimum reference temperature for reparameterization of the Arrhenius equation. Part 1: Problems involving one kinetic constant. *Chem. Eng. Sci.* **2007**, *63*, 2750–2764.
65. Biegler, L.; Grossmann, I.; Westerberg, A. *Systematic Methods for Chemical Process Design*; Prentice Hall: Old Tappan, NJ, USA, 1997.

66. Cappuyns, A.M.; Bernaerts, K.; Smets, I.Y.; Ona, O.; Prinsen, E.; Vanderleyden, J.; Van Impe, J.F. Optimal fed batch experiment design for estimation of monod kinetics of *Azospirillum brasilense*: From theory to practice. *Biotechnol. Prog.* **2007**, *23*, 1074–1081.
67. Telen, D.; Vercammen, D.; Logist, F.; Van Impe, J. Robustifying optimal experiment design for nonlinear, dynamic (bio)chemical systems. *Comput. Chem. Eng.* **2014**, *71*, 415–425.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).